PLEASE AMEND THE SPECIFICATION ACCORDING TO THE FOLLOWING:

Please replace the first paragraph of the specification following the heading CROSS REFERENCE TO RELATED APPLICATIONS with the following amended paragraph:

--The present application is a continuation-in-part of U.S. Patent 6,221,587 filed May 12, 1998, which claims priority to provisional U.S. Serial No. 60/085,092 filed May 12, 1998, each of which is incorporated herein by reference in its entirety.--

Please replace the five paragraphs beginning on page 16, immediately following Table 1, and ending on page 18 of the original specification as filed with the following rewritten paragraphs:

--Additional nucleic acid targets may be determined independently or can be selected from publicly available prokaryotic and eukaryotic genetic databases known to those skilled in the art. Preferred databases include, for example, Online Mendelian Inheritance in Man (OMIM), the Cancer Genome Anatomy Project (CGAP), GenBank, EMBL, PIR, SWISS-PROT, and the like. OMIM, which is a database of genetic mutations associated with disease, was developed, in part, for the National Center for Biotechnology Information (NCBI). OMIM can be accessed through the Internet at, for example, www.ncbi.nlm.nih.gov/Omim/. CGAP, which is an interdisciplinary program to establish the information and technological tools required to decipher the molecular anatomy of a cancer cell. CGAP can be accessed through the Internet at, for example, www.ncbi.nlm.nih.gov/ncicgap/. Some of these databases may contain complete or partial nucleotide sequences. In addition, nucleic acid targets can also be selected from private genetic databases. Alternatively, nucleic acid targets can be selected from available publications or can be determined especially for use in connection with the present invention.

After a nucleic acid target is selected or provided, the nucleotide sequence of the nucleic acid target is determined and then compared to the nucleotide sequences of a plurality of nucleic acids from different taxonomic species. In one embodiment of the invention, the nucleotide sequence of the nucleic acid target is determined by scanning at least one genetic database or is identified in available publications. Preferred databases known and available to those skilled in the art include, for example, the Expressed Gene Anatomy Database

(EGAD) and Unigene-Homo Sapiens database (Unigene), GenBank, and the like. EGAD contains a non-redundant set of human transcript (HT) sequences and can be accessed through the Internet at, for example, www.tigr.org/tdb/egad/egad.html. Unigene is a system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters. Each Unigene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location.

In addition, Unigene contains hundreds of thousands of novel expressed sequence tag (EST) sequences. Unigene can be accessed through the Internet at, for example, www.ncbi.nlm.nih.gov/UniGene/. These databases can be used in connection with searching programs such as, for example, Entrez, which is known and available to those skilled in the art, and the like. Entrez can be accessed through the Internet at, for example, www.ncbi.nlm.nih.gov/Entrez/. Preferably, the most complete nucleic acid sequence representation available from various databases is used. The GenBank database, which is known and available to those skilled in the art, can also be used to obtain the most complete nucleotide sequence. GenBank is the NIH genetic sequence database and is an annotated collection of all publicly available DNA sequences. GenBank is described in, for example, *Nuc. Acids Res.*, 1998, *26*, 1-7, which is incorporated herein by reference in its entirety, and can be accessed by those skilled in the art through the Internet at, for example, www.ncbi.nlm.nih.gov/Web/Genbank/index.html. Alternatively, partial nucleotide sequences of nucleic acid targets can be used when a complete nucleotide sequence is not available.

In another embodiment of the present invention, the nucleotide sequence of the nucleic acid target is determined by assembling a plurality of overlapping expressed sequence tags (ESTs). The EST database (dbEST), which is known and available to those skilled in the art, comprises approximately one million different human mRNA sequences comprising from about 500 to 1000 nucleotides, and various numbers of ESTs from a number of different organisms. dbEST can be accessed through the Internet at, for example, www.ncbi.nlm.nih.gov/dbEST/index.html. These sequences are derived from a cloning strategy that uses cDNA expression clones for genome sequencing. ESTs have applications in the discovery of new genes, mapping of genomes, and identification of coding regions in genomic sequences. Another important feature of EST sequence information that is becoming

rapidly available is tissue-specific gene expression data. This can be extremely useful in targeting selective gene(s) for therapeutic intervention. Since EST sequences are relatively short, they must be assembled in order to provide a complete sequence. Because every available clone is sequenced, it results in a number of overlapping regions being reported in the database.

Assembly of overlapping ESTs extended along both the 5' and 3' directions results in a full-length "virtual transcript." The resultant virtual transcript may represent an already characterized nucleic acid or may be a novel nucleic acid with no known biological function. The Institute for Genomic Research (TIGR) Human Genome Index (HGI) database, which is known and available to those skilled in the art, contains a list of human transcripts. TIGR can be accessed through the Internet at, for example, www.tigr.org/. The transcripts were generated in this manner using TIGR-Assembler, an engine to build virtual transcripts and which is known and available to those skilled in the art. TIGR-Assembler is a tool for assembling large sets of overlapping sequence data such as ESTs, BACs, or small genomes, and can be used to assemble eukaryotic or prokaryotic sequences. TIGR-Assembler is described in, for example, Sutton, *et al.*, *Genome Science & Tech.*, 1995, *1*, 9-19, which is incorporated herein by reference in its entirety, and can be accessed through the Internet at, for example, ftp.tigr.org/pub/software/TIGR assembler. In addition, GLAXO-MRC, which is known and available to those skilled in the art, is another protocol for constructing virtual transcripts. In addition, "Find Neighbors and Assemble EST Blast" protocol, which runs on a UNIX platform, has been developed by Applicants to construct virtual transcripts. Preferred steps in the Find Neighbors and Assemble EST Blast protocol is described in the flowchart set forth in Figure 2. PHRAP is used for sequence assembly within Find Neighbors and Assemble EST Blast. PHRAP can be accessed through the Internet at, for example, chimera.biotech.washington.edu/uwgc/tools/phrap.htm. One skilled in the art can construct source code to carry out the preferred steps set forth in Figure 2.

Please replace the following paragraph beginning on page 19, line 22 and ending on page 20 of the original specification as filed with the following rewritten paragraph:

--Sequence similarity searches can be performed manually or by using several available computer programs known to those skilled in the art. Preferably, Blast and Smith-Waterman algorithms, which are available and known to those skilled in the art, and the like can be used. Blast is NCBI's sequence similarity search tool designed to support analysis of nucleotide and protein sequence databases. Blast can be accessed through the Internet at, for example, www.ncbi.nlm.nih.gov/BLAST/. The GCG Package provides a local version of Blast that can be used either with public domain databases or with any locally available searchable database. GCG Package v.9.0 is a commercially available software package that contains over 100 interrelated software programs that enables analysis of sequences by editing, mapping, comparing and aligning them. Other programs included in the GCG Package include, for example, programs which facilitate RNA secondary structure predictions, nucleic acid fragment assembly, and evolutionary analysis. In addition, the most prominent genetic databases (GenBank, EMBL, PIR, and SWISS-PROT) are distributed along with the GCG Package and are fully accessible with the database searching and manipulation programs. GCG can be accessed through the Internet at, for example, www.gcg.com/. Fetch is a tool available in GCG that can get annotated GenBank records based on accession numbers and is similar to Entrez. Another sequence similarity search can be performed with GeneWorld and GeneThesaurus from Pangea. GeneWorld 2.5 is an automated, flexible, high-throughput application for analysis of polynucleotide and protein sequences. GeneWorld allows for automatic analysis and annotations of sequences. Like GCG, GeneWorld incorporates several tools for homology searching, gene finding, multiple sequence alignment, secondary structure prediction, and motif identification. GeneThesaurus 1.0tm is a sequence and annotation data subscription service providing information from multiple sources, providing a relational data model for public and local data.—

Please replace the following paragraph beginning on page 20, line 23 and ending on page 21 of the original specification as filed with the following rewritten paragraph:

--Another toolkit capable of doing sequence similarity searching and data manipulation is SEALS, also from NCBI. This tool set is written in perl and C and can run on any computer platform that supports these languages. It is available for download, for

example, at: www.ncbi.nlm.nih.gov/Walker/SEALS/. This toolkit provides access to Blast2 or gapped blast. It also includes a tool called tax_collector which, in conjunction with a tool called tax_break, parses the output of Blast2 and returns the identifier of the sequence most homologous to the query sequence for each species present. Another useful tool is feature2fasta which extracts sequence fragments from an input sequence based on the annotation. An exemplary use for this tool is to create sequence files containing the 5′ untranslated region of a cDNA sequence.--

Please replace the following paragraph beginning on page 21, line 29 and ending on page 22 of the original specification as filed with the following rewritten paragraph:

--In another embodiment of the invention, the sequences required are obtained by searching ortholog databases. One such database is Hovergen, which is a curated database of vertebrate orthologs. Ortholog sets may be exported from this database and used as is, or used as seeds for further sequence similarity searches as described above. Further searches may be desired, for example, to find invertebrate orthologs. Hovergen can be downloaded, for example, at: pbil.univ-lyon1.fr/pub/hovergen/. A database of prokaryotic orthologs, COGS, is available and can be used interactively on the internet, for example at: www.ncbi.nlm.nih.gov/COG/.--

Please replace the following paragraph on page 24, lines 15-27 of the original specification as filed with the following rewritten paragraph:

In one embodiment of the invention, secondary structure analysis is performed by alignment and covariance analysis. Numerous protocols for alignment and covariance analysis are known to those skilled in the art. Preferably, alignment is performed by ClustalW, which is available and known to those skilled in the art. ClustalW is a tool for multiple sequence alignment that, although not a part of GCG, can be added as an extension of the existing GCG tool set and used with local sequences. ClustalW can be accessed through the Internet at, for example, dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html. ClustalW is also described in Thompson, *et al.*, *Nuc. Acids Res.*, 1994, *22*, 4673-4680, which is incorporated herein by reference in its entirety. These processes can be scripted to automatically use conserved UTR regions identified in earlier

steps. Seqcd, a UNIX command line interface available and known to those skilled in the art, allows extraction of selected local regions from a larger sequence. Multiple sequences from many different species can be clustered and aligned for further analysis.--

Please replace the following two paragraphs beginning on line 9 of page 25 and ending on page 26 of the original specification as filed with the following rewritten paragraphs:

Covariation is a process of using phylogenetic analysis of primary sequence information for consensus secondary structure prediction. Covariation is described in the following references, each of which is incorporated herein by reference in their entirety: Gutell, *et al.*, "Comparative Sequence Analysis Of Experiments Performed During Evolution" In Ribosomal RNA Group I Introns, Green, Ed., Austin:Landes, 1996; Gautheret, *et al.*, *Nuc. Acids Res.*, 1997, *25*, 1559-1564; Gautheret, *et al.*, *RNA*, 1995, *1*, 807-814; Lodmell, *et al.*, *Proc. Natl. Acad. Sci. USA*, 1995, *92*, 10555-10559; Gautheret, *et al.*, *J. Mol. Biol.*, 1995, *248*, 27-43; Gutell, *Nuc. Acids Res.*, 1994, *22*, 3502-3517; Gutell, *Nuc. Acids Res.*, 1993, *21*, 3055-3074; Gutell, *Nuc. Acids Res.*, 1993, *21*, 3051-3054; Woese, *Proc. Natl. Acad. Sci. USA*, 1989, *86*, 3119-3122; and Woese, *et al.*, *Nuc. Acids Res.*, 1980, *8*, 2275-2293. Preferably, covariance software is used for covariance analysis. Preferably, Covariation, a set of programs for the comparative analysis of RNA structure from sequence alignments, is used. Covariation uses phylogenetic analysis of primary sequence information for consensus secondary structure prediction. Covariation can be obtained through the Internet at, for example, www.mbio.ncsu.edu/RNaseP/info/programs/programs.html. A complete description of a version of the program has been published (Brown, J. W. 1991 Phylogenetic analysis of RNA structure on the Macintosh computer. CABIOS7:391-393). The current version is v4.1, which can perform various types of covariation analysis from RNA sequence alignments, including standard covariation analysis, the identification of compensatory base-changes, and mutual information analysis. The program is well-documented and comes with extensive example files. It is compiled as a stand-alone program; it does not require Hypercard (although a much smaller 'stack' version is included). This program will run in any Macintosh environment running MacOS v7.1 or higher. Faster processor machines (68040 or PowerPC) is suggested for mutual information analysis or the analysis of large sequence alignments.

In another embodiment of the invention, secondary structure analysis is performed by secondary structure prediction. There are a number of algorithms that predict RNA secondary structures based on thermodynamic parameters and energy calculations. Preferably, secondary structure prediction is performed using either M-fold or RNA Structure 2.52. M-fold can be accessed through the Internet at, for example, ~~http://~~www.ibc.wustl.edu/~zuker/rna/form2.cgi or can be downloaded for local use on UNIX platforms. M-fold is also available as a part of GCG package. RNA Structure 2.52 is a windows adaptation of the M-fold algorithm and can be accessed through the Internet at, for example, ~~http://~~128.151.176.70/RNAstructure.html.--

Please replace the following paragraph on page 29, lines 6-26 of the original specification as filed with the following rewritten paragraph:

In one embodiment of the invention, nucleic acids having secondary structure which correspond to the structure descriptor elements are identified by searching at least one database. Any genetic database can be searched. Preferably, the database is a UTR database, which is a compilation of the untranslated regions in messenger RNAs. A UTR database is accessible through the Internet at, for example, area.ba.cnr.it/pub/embnet/database/utr/. Preferably the database is searched using a computer program, such as, for example, Rnamot, a UNIX-based motif searching tool available from Daniel Gautheret. Each "new" sequence that has the same motif is then queried against public domain databases to identify additional sequences. Results are analyzed for recurrence of pattern in UTRs of these additional ortholog sequences, as described below, and a database of RNA secondary structures is built. One skilled in the art is familiar with Rnamot. Briefly, Rnamot takes a descriptor string, such as the one shown in Figure 9, and searches any Fasta format database for possible matches. Descriptors can be very specific, to match exact nucleotide(s), or can have built-in degeneracy. Lengths of the stem and loop can also be specified. Single stranded loop regions can have a variable length. G-U pairings are allowed and can be specified as a wobble parameter. Allowable mismatches can also be included in the descriptor definition. Functional significance is assigned to the motifs if their biological role is known based on previous analysis. Known regulatory regions such as Iron Response Element have been found using this technique (see, Example 1 below). In embodiments of the invention in which a database

containing prokaryotic molecular interaction sites is compiled, it is preferable to refrain from searching human sequences or, alternatively, discarding human sequences when found.--